BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 5: e74359
doi: 10.3897/biss.5.74359

OPEN ACCESS

Conference Abstract

Research Infrastructure Contact Zones: A method to visualise and align the activities of major biodiversity informatics initiatives

Vincent Stuart Smith[‡], Aino Juslén[§], Ana Casino[|], Francois Dusoulier[¶], Lisa French[‡], Dimitrios Koureas[#,¤], Patricia Mergen[«,»], Joe Miller[^], Leif Schulman[˅], Matt Woodburn[‡]

‡ The Natural History Museum, London, United Kingdom
§ Finnish Museum of Natural History, Helsinki, Finland
| CETAF, Brussels, Belgium
¶ Muséum national d'Histoire naturelle, Paris, France
# Naturalis Biodiversity Center, Leiden, Netherlands
¤ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
« Meise Botanic Garden, Meise, Belgium
» Royal Museum for Central Africa, Tervuren, Belgium
^ GBIF, Copenhagen, Denmark
˅ Finnish Environment Institute, Helsinki, Finland

Corresponding author: Vincent Stuart Smith (v.smith@nhm.ac.uk)

Received: 13 Sep 2021 | Published: 13 Sep 2021

Citation: Smith VS, Juslén A, Casino A, Dusoulier F, French L, Koureas D, Mergen P, Miller J, Schulman L, Woodburn M (2021) Research Infrastructure Contact Zones: A method to visualise and align the activities of major biodiversity informatics initiatives. Biodiversity Information Science and Standards 5: e74359.
https://doi.org/10.3897/biss.5.74359

Abstract

In an effort to characterise the various dimensions of activity within the biodiversity informatics landscape, we developed a framework to survey these dimensions for ten major organisations*[1] relative to both their current activities and long-term strategic ambitions. This survey assessed the contact between these infrastructure organisations by capturing the breadth of activities for each infrastructure across five categories (data, standards, software, hardware and policy), for nine types of data (specimens, collection descriptions, opportunistic observations, systematic observations, taxonomies, traits, geological data, molecular data, and literature), and for seven phases of activity (creation, aggregation, access, annotation, interlinkage, analysis, and synthesis). This generated a dataset of 6,300 verified observations, which have been scored and validated by leading

© Smith V et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

members of each infrastructure organisation. In this analysis of the resulting data, we address a set of high-level questions about the overall biodiversity informatics landscape, looking at the greatest gaps, overlap and possible rate-limiting steps. Across the infrastructure organisations, we also explore how far each is in relation to achieving its ambitions and the extent of its niche relative to other organisations.

Our results show that when viewed by scope, most infrastructures occupy a relatively narrow niche in the overall landscape of activity, with the notable exception of the Global Biodiversity Information Facility (GBIF) and possibly LifeWatch. Niches associated with molecular data and biological taxonomy are very well filled, suggesting there is still considerable room for growth in other areas, with the Distributed System of Scientific Collections (DiSSCo) and the Integrated European Long-Term Ecosystem Research Infrastructure (eLTER RI) showing the highest levels of difference between their current activities and stated ambitions, potentially reflecting the relative youth of these organisations. iNaturalist, the Biodiversity Heritage Library and Catalogue of Life all occupy narrow and tightly circumscribed niches. These organisations are also amongst the closest to achieving their stated ambitions within their respective areas of activity. The largest gaps in infrastructure activity relate to the development of hardware and standards, with many gaps set to be addressed if the stated ambitions of those surveyed come to fruition. Nevertheless, some gaps persist, outlining a potential role for this survey as a planning tool to help coordinate and align investment in future biodiversity informatics activities. GBIF and LifeWatch are the two infrastructures where there is the most similarity in ambition with DiSSCo, with the greatest overlap concentrated on activities related to data/content, specimen data and their shared ambition to interlink information. While overlap appears intense, the analysis is limited by the resolution of the survey framework and ignores existing collaborations between infrastructures.

In addition to presenting the results of this survey, we outline our plans to publish this work and a proposal to develop the methodology as an interactive web-based tool. This would allow other projects and infrastructures to self-score their activities and visualise their niche within the current landscape, encouraging better global alignment of activities. For example, our results should make it easier for initiatives to strengthen collaboration and differentiate work when their activities overlap. Likewise, this approach would be useful for funding agencies when targeting gaps in the informatics landscape or increasing the technical maturity of certain critical activities, e.g., to improve immature data standards. While no framework is perfect, we hope to encourage a dialogue on the potential for taking an algorithmic approach to community alignment and see this as a means of strengthening community cooperation when addressing problems that require global cooperation.

Keywords

methods, data visualisation, coordination, alignment, community

Presenting author

Vincent S. Smith

Presented at

TDWG 2021

Acknowledgements

We wish to thank DiSSCo, GBIF, iBOL, the Catalogue of Life, iNaturalist, BHL, GeoCASe, LifeWatch, eLTER RI and ELIXIR for the considerable time and effort they spent to provide the data that underpins this analysis. Their comments also considerably improved the technical definitions and methods used in this study.

Endnotes

*1 The organisations surveyed were the Distributed System of Scientific Collections (DiSSCo), the Global Biodiversity Information Facility (GBIF), the International Barcode of Life (iBOL), the Catalogue of Life, iNaturalist, the Biodiversity Heritage Library, GeoCASe, LifeWatch, the integrated European Long-Term Ecosystem, critical zone and socio-ecological Research Infrastructure (eLTER), and ELIXIR.